W1526

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-203201
(43)Date of publication of application : 30.07.1999

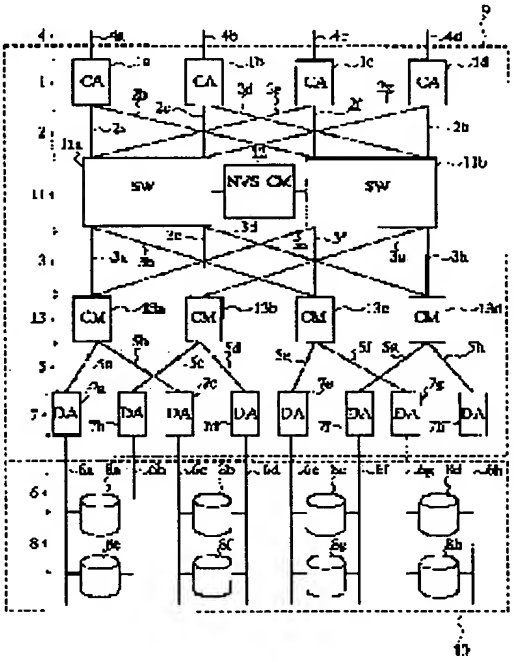
(51)Int.Cl. G06F 12/08
G06F 12/08

(21)Application number : 10-002400
(22)Date of filing : 08.01.1998

(71)Applicant : HITACHI LTD
(72)Inventor : MORI KENJI

(54) ARRANGING METHOD OF CACHE MEMORY AND DATA STORAGE SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To dissolve bottleneck of an access to a cache memory caused by an increase of the number of paths or storage devices of a host device.
SOLUTION: A non-volatile cache memory (12) where write-in data from the host device are stored and a volatile cache memory 143 where data read from a disk drive 8 are temporarily stored are separately provided between plural channel I/F control circuits 1 that control plural channels I/F 4 on the side of the host device and plural disk control circuits 7 that control plural disk drives 8 in a disk drive unit 10. Moreover, the non-volatile cache memory (12) is commonly and concentratedly provided in plural data transmission paths, the volatile cache memory 13 is distributed every several data transmission paths and is arranged, and set of each capacity or throughput of the non-volatile cache memory (12) and plural volatile cache memories 13 is individually enabled.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号

特開平11－203201

(43)公開日　平成11年(1999)７月30日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　12/08	３２０	Ｇ０６Ｆ　12/08	Ｇ ３２０

審査請求　未請求　請求項の数３　ＯＬ　（全 13 頁）

(21)出願番号　特願平10－2400

(22)出願日　平成10年(1998)１月８日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　森　健治
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(74)代理人　弁理士　筒井　大和

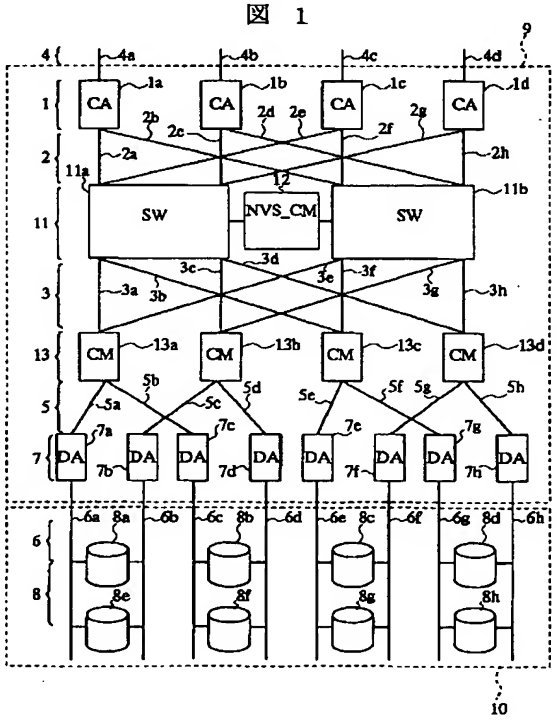
(54)【発明の名称】　キャッシュメモリの配置方法およびデータ記憶システム

(57)【要約】

【課題】　上位装置側のパス数や記憶装置数の増大に起因するキャッシュメモリへのアクセスのボトルネックを解消する。

【解決手段】　上位装置側の複数のチャネルＩ／Ｆ４を制御する複数のチャネルＩ／Ｆ制御回路１と、ディスクドライブユニット１０における複数のディスクドライブ８を制御する複数のディスク制御回路７との間に、上位装置側からの書き込みデータが格納される不揮発キャッシュメモリ１２と、ディスクドライブ８から読出されたデータが一時的に格納される揮発キャッシュメモリ１３を別個に設置するとともに、不揮発キャッシュメモリ１２は、複数のデータ転送経路に共通に集中して設置し、揮発キャッシュメモリ１３は、いくつかのデータ転送経路毎に分散して配置し、不揮発キャッシュメモリ１２および複数の揮発キャッシュメモリ１３の各々の容量やスループットを個別に設定可能にした。

図　1



【特許請求の範囲】
【請求項１】　上位装置と、前記上位装置との間で授受される情報が格納される記憶装置との間に配置され、前記情報を一時的に保持するキャッシュメモリの配置方法であって、
前記キャッシュメモリを不揮発キャッシュメモリおよび揮発キャッシュメモリにて構成し、前記不揮発キャッシュメモリは集中的に配置し、前記揮発キャッシュメモリは分散して配置することを特徴とするキャッシュメモリの配置方法。
【請求項２】　上位装置と、前記上位装置との間で授受される情報が格納される記憶装置との間に配置され、前記情報を一時的に保持するキャッシュメモリの配置方法であって、
前記キャッシュメモリを不揮発キャッシュメモリおよび揮発キャッシュメモリにて構成し、
前記不揮発キャッシュメモリの容量と、前記揮発キャッシュメモリの容量とが異なるように設定する第１の方法、
前記不揮発キャッシュメモリのスループットと、前記揮発キャッシュメモリのスループットとが異なるように設定する第２の方法、
の少なくとも一方の方法を用いることを特徴とするキャッシュメモリの配置方法。
【請求項３】　上位装置との間で授受される情報が格納される記憶装置と、前記記憶装置と前記上位装置との間における前記情報の授受を制御する記憶制御装置と、前記上位装置と前記記憶装置との間に配置され、前記情報が一時的に格納されるキャッシュメモリとを含むデータ記憶システムであって、
前記キャッシュメモリは、不揮発キャッシュメモリおよび揮発キャッシュメモリからなり、
前記不揮発キャッシュメモリは集中的に配置され、前記揮発キャッシュメモリは分散して配置される第１の構成、
前記不揮発キャッシュメモリの容量と、前記揮発キャッシュメモリの容量とが異なる第２の構成、
前記不揮発キャッシュメモリのスループットと、前記揮発キャッシュメモリのスループットとが異なる第３の構成、
の少なくとも一つの構成を備えたことを特徴とするデータ記憶システム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、キャッシュメモリの配置技術およびデータ記憶技術に関し、特に、上位装置と記憶装置との間に設けられた複数の情報転送経路にて並列的に情報の授受が行われるデータ記憶システム等に適用して有効な技術に関する。
【０００２】
【従来の技術】図１３は、考えられる従来のディスク制御装置１０９および配下のディスクドライブユニット１１０からなるディスクサブシステムの構成の一例を示す概念図である。従来のディスク制御装置１０９では、ホスト側の入出力制御を行うチャネル制御回路１０１と、ディスクドライブ側の入出力制御を行うディスク制御回路１０７との間に、キャッシュメモリアクセスパス１０５およびキャッシュメモリアクセスパス１０６を介してアクセスされるキャッシュメモリを配置する構成が一般的である。このとき、キャッシュメモリは、電源障害に起因するデータ喪失に備えた不揮発キャッシュメモリ１０２、またキャッシュメモリの障害時の補償用としての揮発キャッシュメモリ１０３、を持ち単純な２重系として使用していた。この場合、装置の対ホストに対するパス１０４や、ディスクドライブ１０８の数が増加した場合、全データアクセスにおいて２つの不揮発キャッシュメモリ１０２および揮発キャッシュメモリ１０３にアクセスが発生し、キャッシュメモリにアクセス集中するため、ディスク制御装置１０９の性能におけるボトルネックとなる。
【０００３】しかし従来のディスク制御装置１０９では対ホストに対するパス１０４、ディスクドライブ１０８の数が図１３の様に４チャネル、８ドライブパス程度と少なく、この様な方式である程度の性能を確保できるため、この様な方式が取られてきた。
【０００４】
【発明が解決しようとする課題】今後ディスク制御装置当たりの容量の増加、高機能化により、制御を行うディスクドライブの数が多くなり、また装置のホストに対するパス数が増加すると考えられる。
【０００５】この様な制御装置を従来の方法で構成した場合、キャッシュメモリへのアクセスが集中し、キャッシュメモリが制御装置のボトルネックとなる。このボトルネックを解消するためには、キャッシュメモリを分散配置する方法が考えられる。この時、従来の様にそれぞれ障害に備え不揮発キャッシュメモリを含む２重系のキャッシュメモリを用意する必要があり、この様なもので分散キャッシュを構成すると実装的に困難で、また価格的にも高価となる。
【０００６】本発明の目的は、キャッシュメモリを経由したデータ転送のスループットを向上させることが可能なキャッシュメモリの配置技術およびデータ記憶技術を提供することにある。
【０００７】本発明の他の目的は、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおけるコスト削減を実現することが可能なキャッシュメモリの配置技術およびデータ記憶技術を提供することにある。
【０００８】本発明の他の目的は、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッ

シュメモリにおける実装効率の向上を実現することが可能なキャッシュメモリの配置技術およびデータ記憶技術を提供することにある。

【０００９】

【課題を解決するための手段】本発明では、ライト時のデータを保証するための不揮発キャッシュメモリを１箇所に配置する。またリードの効率を上げるために、複数のアクセスができる様に揮発キャッシュメモリを分散配置する。この分散配置される揮発キャッシュメモリはリード時のみに使用し、データが万一電源切断等で揮発してもディスクからリードできるため、揮発性のメモリ媒体で構成可能である。

【００１０】また、ディスク制御装置等の記憶制御装置ではリード／ライトの比率は対称でなく、リード４に対しライト１程度で使用される場合が多い。この性質を利用するとメモリ容量はリード／ライトの比率により不揮発キャッシュメモリはリード用の揮発キャッシュメモリほど大きくしなくても、性能に対して影響しない。またスループットも不揮発キャッシュメモリはリード用の揮発キャッシュメモリより小さくできる。しかし分散配置するリード用の揮発キャッシュメモリの容量は敏感に性能に対して影響するため、不揮発キャッシュメモリよりも容量を大きくする必要がある。この様に、キャッシュメモリの配置を分散と集中の２つの組で、それぞれの容量を要求される性能に応じた値に設定することでキャッシュメモリを最適量にすることができ、またこれにより実装的、価格的にも有利に、低価格で高性能化を図ることができる。

【００１１】今後、ディスク制御装置当たりの容量の増加、高機能化により、制御を行うディスクドライブの数が多くなり、また装置のホストに対するパスが増加すると考えられる。

【００１２】その場合、本発明では、複数のパスに対して不揮発キャッシュメモリを１箇所に集中的に配置し、揮発キャッシュメモリを、たとえば、一つあるいは幾つかのパス毎に分散配置する。これら揮発キャッシュメモリを分散することにより、同時に対キャッシュメモリ転送が可能となり、多チャネル、多ドライブ化への対応時に１つのキャッシュメモリに集中することなく高いスループットをもたらすことができる。また分散するキャッシュメモリは揮発性でよいため実装的にも有利で、また低価格で構成可能である。さらに増設単位を分散配置される揮発キャッシュメモリのサイズを単位とすることにより、自由なスループット構成を採ることが可能となる。この様なキャッシュメモリの構成とすることで、たとえば複数パス構成のシステムにおいて低価格で高性能化を図ることが可能となる。

【００１３】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【００１４】図１は本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの構成の一例を示す概念図である。本実施の形態では、データ記憶サブシステムの一例として、ディスクサブシステムに適用した場合を例に採って説明する。

【００１５】図１に例示される本実施の形態のディスクサブシステムは大きく、ディスク制御ユニット９、ディスクドライブユニット１０、の２つのユニットより構成されている。ディスク制御ユニット９は、複数のチャネルＩ／Ｆ制御回路１（１ａ～１ｄ）、複数のデータパススイッチ１１（１１ａ～１１ｂ）、不揮発キャッシュメモリ１２、複数の揮発キャッシュメモリ１３（１３ａ～１３ｄ）、複数のディスク制御回路７（７ａ～７ｈ）、より構成され、ディスクドライブユニット１０は複数のディスクドライブ８（８ａ～８ｈ）より構成されている。

【００１６】図１において、複数のチャネルＩ／Ｆ制御回路１ａ～１ｄの各々は、複数のチャネルＩ／Ｆ４（４ａ～４ｄ）を介して図示しないチャネル装置や中央処理装置（ＣＰＵ）等のホスト装置に個別に接続され、チャネルＩ／Ｆ４のプロトコル制御、データ変換、データ転送を行う。

【００１７】複数のデータパススイッチ１１は、複数のパス２（２ａ～２ｈ）および複数のパス３（３ａ～３ｈ）を介して接続されている複数のチャネルＩ／Ｆ制御回路１と、不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３との間におけるデータの流れの経路を切り替える役目をはたす。

【００１８】複数の揮発キャッシュメモリ１３は、複数のパス５（５ａ～５ｈ）を介して複数のディスク制御回路７に接続され、複数のディスク制御回路７は、複数のパス６（６ａ～６ｈ）を介してディスクドライブユニット１０に接続されている。

【００１９】不揮発キャッシュメモリ１２は、ホスト側からのライト時のデータを格納するために用いる。これはライト処理の高速化の目的で、ホスト側からのライトコマンドに対して、不揮発キャッシュメモリ１２への書き込みが完了した時点で、ディスクドライブ８に書き込む前に、完了を返す制御を行うためで、万一、ディスクドライブ８にデータを書き込む前に停電で装置が停止した際のデータを、この不揮発キャッシュメモリ１２でバックアップする目的に使用する。また、この不揮発キャッシュメモリ１２に書き込む時は、可能な限り、同時に揮発キャッシュメモリ１３に書き込みを行うようにデータパススイッチ１１で経路の設定を行う。

【００２０】揮発キャッシュメモリ１３は主にリードデータのキャッシュメモリとして働き、ディスクドライブ８からデータをリードした際に、この揮発キャッシュメモリ１３にデータを格納しておき、２回目以降に同じデータに対するリードに対してはディスクドライブ８から

データをリードするのではなく、揮発キャッシュメモリ１３からリードすることにより高速化を図る目的に用いる。このためこの揮発キャッシュメモリ１３は停電等により記憶データが揮発した場合でもディスクドライブにデータが保存されているため、揮発性のメモリ媒体で構成する事が可能である。

【００２１】ディスク制御回路７は、パス６を介してディスクドライブ８に接続されており、ディスクドライブ８の制御を行うのに用いる。１つのディスク制御回路７は２つのディスクドライブ制御用のパスを有し、これにより万一の障害発生時にも交代パスを使用することによりディスクドライブ８へのアクセスが可能である。また通常時は２パスを有効に使用することで高速化の機能もはたす。

【００２２】ディスクドライブ８は、データを格納するのに使用され、ディスクドライブ８へのアクセスパスは高速化、高信頼化のための２つの経路を有する。

【００２３】すなわち、本実施の形態の場合には、ディスクドライブユニット１０を構成する複数のディスクドライブ８は、いくつかのグループ（本実施の形態では一例として４つ）に系列化され、各系列は、複数のパス６ａ～６ｂ、パス６ｃ～６ｄ、パス６ｅ～６ｆ、パス６ｇ～６ｈ、をそれぞれ介して、多重経路でディスク制御ユニット９に接続されている。また、複数のパス６ａ～６ｈの各々におけるデータ転送は、複数のディスク制御回路７ａ～７ｈの各々にて独立に制御される。

【００２４】以下、本実施の形態のキャッシュメモリの配置方法およびディスクサブシステムの作用の一例について説明する。

【００２５】図２はＣＰＵからのリードコマンドの実行時のデータフローの一例を示した概念図である。この図２では、たとえばチャネルＩ／Ｆ制御回路１ａにデータリードコマンドがチャネルＩ／Ｆ４ａを介してＣＰＵから入った場合について説明する。

【００２６】まず、チャネルＩ／Ｆ制御回路１ａでＣＰＵからのコマンドを認識する。この結果リードコマンドであることがわかる。その後、どのディスクドライブ８のデータかを判別し、判別結果に応じて、たとえば、データパススイッチ１１ａにおいて切り替え、チャネルＩ／Ｆ制御回路１ａと揮発キャッシュメモリ１３ａとを結合状態とする。

【００２７】次にデータがキャッシュメモリ上にあるかの判定を行う。この判定はチャネルＩ／Ｆ制御回路１ａで行う。この結果、対象データがキャッシュメモリ上にある（キャッシュヒット）と判定した場合には、データフロー１６の様に揮発キャッシュメモリ１３ａよりチャネル制御回路１ａにデータをデータパススイッチ１１ａ通じて転送することで読み出し、チャネルＩ／Ｆ制御回路１ａがチャネルＩ／Ｆ４ａを介してホスト（ＣＰＵ）にデータを転送し、リードコマンドが終了する。

【００２８】もしアクセス対象データが揮発キャッシュメモリ１３ａ上に無い（キャッシュミス）場合は、データフロー１７の様にディスク制御回路７ａ経由でディスクドライブ８ａのリードを行い、データをデータパススイッチ１１ａを通じて転送し読み出す。この際、同時にデータを揮発キャッシュメモリ１３ａにライトする。これにより２回目以降のデータリードに対してはキャッシュヒットとなりアクセスの高速化を図ることができる。

【００２９】図３はＣＰＵからのライトコマンドの実行時のデータフローの一例を示した概念図である。チャネルＩ／Ｆ制御回路１ａにチャネルＩ／Ｆ４ａを介してデータライトコマンドがＣＰＵから入った場合について説明する。

【００３０】まずチャネルＩ／Ｆ制御回路１ａでＣＰＵからのコマンドを認識する。この結果ライトコマンドであることがわかる。この後ライト対象がどのディスクドライブ８かを判別し、判別結果に応じてデータパススイッチ１１ａを切り替え、たとえば、チャネルＩ／Ｆ制御回路１ａと不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３ａを結合状態とする。

【００３１】チャネルＩ／Ｆ制御回路１ａからのデータをデータフロー２１の様に同時に揮発キャッシュメモリ１３ａ、不揮発キャッシュメモリ１２に転送する。これによりライトデータは不揮発キャッシュメモリ１２上にライトされるとともに、分散キャッシュメモリ１３ａに転送される。この後にパス５ａを介してディスク制御回路７ａにもライトデータが転送され、さらにパス６ａを介して目的のディスクドライブ８ａに転送されることによって当該ディスクドライブ８ａにデータが書き込まれる。

【００３２】次に、停電が発生した場合の動作について説明する。チャネルＩ／Ｆ制御回路１ａがデータを受け取り、データフロー２２の様に不揮発キャッシュメモリ１２に転送が完了した時点でチャネルＩ／Ｆ制御回路１ａはＣＰＵに対しコマンド完了を返す。このためＣＰＵではデータ書き込みが完了したと判断することになる。しかしこの時点ではディスク制御ユニット９内にデータが在り、まだディスクドライブ８ａに書き込まれていない状態にある。このとき停電が発生すると、転送が完了した不揮発キャッシュメモリ１２と揮発キャッシュメモリ１３ａ上の内、揮発キャッシュメモリ１３ａ上のデータは消えてしまう。この後、電源が復旧した時に、未書き込みのデータが不揮発キャッシュメモリ１２上に存在するかの確認を行い、ディスク制御ユニット９内のデータのディスクドライブ８ａへの書き込みが再開される。この時、データフロー２３の様に不揮発キャッシュメモリ１２からのデータは、データパススイッチ１１ａから、パス３ａ、揮発キャッシュメモリ１３ａ、パス５ａ、ディスク制御回路７ａ、パス６ａ、を通じて目的のディスクドライブ８ａへ転送される。

【００３３】図４は複数のリード／ライト発生時のデータフローの一例を示す概念図である。

【００３４】これはチャネルＩ／Ｆ制御回路１ａにはチャネルＩ／Ｆ４ａを介してライトコマンドの要求が、他のチャネルＩ／Ｆ制御回路１ｂ、１ｃ、１ｄにはチャネルＩ／Ｆ４ｂ、４ｃ、４ｄを介してリードコマンドの要求があった場合の処理の一例を示す図である。

【００３５】まず、チャネルＩ／Ｆ制御回路１ａの動作について説明する。チャネルＩ／Ｆ制御回路１ａはライトコマンドを実行するため、データパススイッチ１１ａを制御して不揮発キャッシュメモリ１２と揮発キャッシュメモリ１３ａへのパスを選択する。そしてＣＰＵからのライトデータはデータフロー３４の様に、パス２ａ、データパススイッチ１１ａを経由して不揮発キャッシュメモリ１２に転送され、同時に、パス２ａ、データパススイッチ１１ａ、パス３ａ、揮発キャッシュメモリ１３ａ、パス５ａを介してディスク制御回路７ａに転送され、さらにパス６ａを経由してディスクドライブ８ａに書き込まれる。

【００３６】次にチャネルＩ／Ｆ制御回路１ｂ、１ｃの動作について説明する。チャネルＩ／Ｆ制御回路１ｂでリードコマンドを実行するため、データパススイッチ１１ａは揮発キャッシュメモリ１３ｂを選択する。この時、データパススイッチ１１ａはチャネルＩ／Ｆ制御回路１ａとチャネルＩ／Ｆ制御回路１ｂの２つの要求に応じ同時に２種類のパスを形成できる。そしてリード対象のデータが揮発キャッシュメモリ１３ｂに存在した場合（キャッシュヒット）、データフロー３５の様に、揮発キャッシュメモリ１３ｂからパス３ｃ、パス２ｃを経由してチャネルＩ／Ｆ制御回路１ｂへデータ転送が行われる。

【００３７】また、チャネルＩ／Ｆ制御回路１ｃではリードコマンドを実行するため、データパススイッチ１１ｂは揮発キャッシュメモリ１３ｃを選択する。そしてリード対象のデータが揮発キャッシュメモリ１３ｃに存在した場合（キャッシュヒット）、データフロー３６の様に、揮発キャッシュメモリ１３ｃから、パス３ｆ、パス２ｆを経由してチャネルＩ／Ｆ制御回路１ｃへデータ転送が行われる。

【００３８】最後にチャネルＩ／Ｆ制御回路１ｄの動作について説明する。チャネルＩ／Ｆ制御回路１ｄでリードコマンドを実行するため、データパススイッチ１１ｂは揮発キャッシュメモリ１３ｄを選択する。この揮発キャッシュメモリ１３ｄに対象データが存在しない場合（キャッシュミス）、データフロー３７の様にパス５ｈを経由してディスク制御回路７ｈにリード要求が入り、パス６ｈを経由してディスクドライブ８ｄから読出されたリードデータは揮発キャッシュメモリ１３ｄにライトされながら、パス３ｈ、パス２ｈを経由してチャネルＩ／Ｆ制御回路１ｄに転送される。この様に揮発キャッシュメモリ１３を、パス毎に１３ａ～１３ｄのように分散配置することにより、複数のパスに対するデータ転送を同時に行うことが可能となる。

【００３９】図５は複数のリード／ライトが特定のディスクドライブ（たとえばディスクドライブ８ｄ）に集中した場合のデータフローの一例を示す概念図である。

【００４０】これはチャネルＩ／Ｆ制御回路１ａ、１ｂにライトコマンド、チャネルＩ／Ｆ制御回路１ｃ、１ｄにリードコマンドの要求があった場合の図である。

【００４１】まずチャネルＩ／Ｆ制御回路１ａの動作について説明する。データフロー４３の様に、チャネルＩ／Ｆ制御回路１ａはライトコマンドを実行するため、データパススイッチ１１ａを不揮発キャッシュメモリ１２と揮発キャッシュメモリ１３ｃへのパスを選択する。しかしこの時、リードコマンド等によりディスク制御回路７ｇが使用されている場合は、不揮発キャッシュメモリ１２までの転送のみを先に行い、その後、ディスク制御回路７ｇが空いてから、停電時と同じ様に不揮発キャッシュメモリ１２から、揮発キャッシュメモリ１３ｃを経由してディスク制御回路７ｇにデータを転送し、ディスクドライブ８ｄに書き込む。図５の例の場合は、ディスク制御回路７ｇが空いているため同時に転送を行う。

【００４２】ホストに対するコマンドの終了報告はディスクドライブ８ｄに書き込みが完了する前に、不揮発キャッシュメモリ１２に書込んだ時点で行う。基本的には不揮発キャッシュメモリ１２へは確実に書き込み、その時、ディスク制御回路７ｇへのパスが空いている場合は同時にディスク制御回路７ｇへの転送を行う処理とする。

【００４３】チャネルＩ／Ｆ制御回路１ｂの動作についても同様に、データフロー４４の様に一時的に不揮発キャッシュメモリ１２に書き込みを行い、その後、ディスク制御回路７ｈが空いてから不揮発キャッシュメモリ１２から、揮発キャッシュメモリ１３ｄを経由してディスク制御回路７ｈにデータを転送し、ディスクドライブ８ｄに書き込む。この例では不揮発キャッシュメモリ１２に対する書き込みは、２つのチャネルＩ／Ｆ制御回路１ａ、１ｂよりのライト要求が競合するがこれは順番に行う。

【００４４】次にチャネルＩ／Ｆ制御回路１ｃの動作について説明する。チャネルＩ／Ｆ制御回路１ｃでリードコマンドを実行するため、データパススイッチ１１ｂが揮発キャッシュメモリ１３ｄを選択する。そしてリード対象のデータが揮発キャッシュメモリ１３ｄに存在した場合（キャッシュヒット）、データフロー４５の様に揮発キャッシュメモリ１３ｄから、パス３ｈ、パス２ｆを経由してチャネルＩ／Ｆ制御回路１ｃへデータ転送が行われる。

【００４５】もし揮発キャッシュメモリ１３ｄに目的のデータが存在しない場合（キャッシュミス）は、パス６

ｈ、ディスク制御回路７ｈ、パス５ｈ、揮発キャッシュメモリ１３ｄ、パス３ｈを経由してディスクドライブ８ｄよりデータを読み出す。

【００４６】チャネルＩ／Ｆ制御回路１ｄについても同様に、データフロー４６の様に、まず揮発キャッシュメモリ１３ｄにリード対象のデータがあるかの確認を行い、ある場合（キャッシュヒット）には揮発キャッシュメモリ１３ｄより転送を行い、無い場合（キャッシュミス）はディスクドライブ８ｄよりデータを読み出す。この時、リードコマンドが競合するが、これは順番に行う。

【００４７】従来の技術では、この様な場合、単一のディスクドライブに対するリード、ライトでも分散したディスクドライブに対するアクセスでも、キャッシュメモリに対するアクセスの競合が発生し、待ちが多くなることになり、性能低下の原因となった。

【００４８】しかし本実施の形態では、分散したディスクドライブに対するアクセスでは競合は発生せず高性能が得られる。また図５の様に１つのディスクドライブ８ｄに対するアクセスが集中するケースは、短いアクセスが多数発生するトランザクション処理では少ないため、リード時の特定の揮発キャッシュメモリに対するアクセスの集中は少ないといえる。また集中管理するメモリとして不揮発キャッシュメモリ１２がありライト時にアクセスが集中するが、ライトの比率が一般的にリードに比べ４分の１程度のため、それほどの問題にはならない。そのため、これにより不揮発キャッシュメモリ１２の容量も他の揮発キャッシュメモリ１３ａ～１３ｄの総和に比べて少なくてすみ、低価格で高性能なディスクサブシステムを実現することが可能となる。

【００４９】ここでキャッシュメモリのサイズ、スループットに関する説明を行う。図５の様に不揮発キャッシュメモリ１２を１つ、また揮発キャッシュメモリ１３を、一例として１３ａ、１３ｂ、１３ｃ、１３ｄの４組み持った構成の場合を考える。ライトとリードに対する比率を１：３とする。これによりに不揮発キャッシュメモリ１２のサイズ、スループットはリードに用いる複数の揮発キャッシュメモリ１３（１３ａ～１３ｄ）の等価サイズ、スループットの４分の１程度ですむことになり、実装的、価格的に都合がよい。

【００５０】次に複数個の揮発キャッシュメモリ１３ａ～１３ｄから構成される揮発キャッシュメモリ１３のサイズ、スループットについて考える。この時、複数の揮発キャッシュメモリ１３ａ～１３ｄの各々のカーバー範囲がディスクドライブパス（パス５ａと５ｂ、パス５ｃと５ｄ、パス５ｅと５ｆ、パス５ｇと５ｈ）とした場合、このディスクドライブパスに対する集中の度合いでその値が決まる。図５の様にパス５ａ～５ｈの８組で構成され、それぞれの競合がないとする場合、等価スループットそのもので良いが、平均２つの競合が発生すると した場合は等価スループットの２倍の性能が必要になってくる。またサイズに関しては、単純には組み数（パス５の数）で割れば良いが、データの分散の程度により値が異なる。一部に集中した場合を想定した場合にはその分多く持つことでキャッシュメモリの効果を引き出すことができる。これは確率的に値を決めることになる。

【００５１】ところで、本実施の形態のように、不揮発キャッシュメモリ１２と、揮発キャッシュメモリ１３（１３ａ～１３ｄ）とを分散して配置し、不揮発キャッシュメモリ１２に対するデータ書き込みが完了した時点でホスト側に書き込み完了を応答する構成では、不揮発キャッシュメモリ１２への書き込みデータが、必ずしも直ちに揮発キャッシュメモリ１３やディスクドライブ８に反映されているとは限らない。このため、未反映の間にリード要求が発生した場合には、最新のデータが、不揮発キャッシュメモリ１２、揮発キャッシュメモリ１３、ディスクドライブ８のいずれに存在するかを判別する操作が必要となる。

【００５２】本実施の形態では、一例として、図６に例示されるような制御情報を用いて、このような判別操作を行う。

【００５３】すなわち、不揮発キャッシュメモリ１２では、たとえばアクセス単位のエントリ毎に、ＮＶＳ管理フラグ５０（V_N）を設ける。本実施の形態の場合、V_Nが“０”のとき、当該エントリの書き込みデータは揮発キャッシュメモリ１３に未反映であり、“１”のときは反映済である。

【００５４】また、揮発キャッシュメモリ１３では、たとえばアクセス単位のエントリ毎に、ＣＭ管理フラグ５１（Ｖ，Ａ）を設ける。本実施の形態の場合、Ｖが“０”のとき、当該エントリのデータに対して、不揮発キャッシュメモリ１２に未反映のデータが存在し、Ｖが“１”のときは存在しない。また、Ａが“０”のとき、当該エントリの書き込みデータはディスクドライブ８上に未反映の状態にあり、Ａが“１”のときは反映済である。

【００５５】なお、ＣＭ管理フラグ５１においては、電源投入直後は、格納データが消失しているが、この状態では、全エントリのＶおよびＡは、ともに“０”の状態にあり、この状態では、キャッシュミスと判定され、ディスクドライブ８上からのデータリードが実行される。そして、不揮発キャッシュメモリ１２に存在するディスクドライブ８に未反映のデータの当該ディスクドライブ８への書き込み操作や、ディスクドライブ８から読出されたデータの格納操作によって、ＶおよびＡは後述のように変化する。

【００５６】そして、データ書き込みに際しては、たとえば、図７のフローチャートに例示されるように、ホスト側（チャネルＩ／Ｆ制御回路１）から到来する書き込みデータを、不揮発キャッシュメモリ１２に書き込んだ

のち、ＮＶＳ管理フラグ５０のV_Nを“０”にセットし（ステップ２０１）、さらにＣＭ管理フラグ５１のＶを“０”にセットする（ステップ２０２）。その後、ホスト側にライト完了を応答する（ステップ２０３）。なお、ステップ２０２ではＣＭ管理フラグ５１のＶの操作のために揮発キャッシュメモリ１３へのアクセスが発生するが通常のデータ転送とは異なり、わずかなフラグビットの操作のみであるため、オーバーヘッドは少ない。

【００５７】たとえば、上述の図５の例のように、不揮発キャッシュメモリ１２以下への書き込みデータの転送は、任意契機でよく、たとえば、図８のフローチャート例示されるような手順にて行われる。

【００５８】すなわち、まず、ＮＶＳ管理フラグ５０のV_Nが“０”のエントリを不揮発キャッシュメモリ１２から検索し（ステップ３０１）、当該データを、揮発キャッシュメモリ１３に転送した後、ＣＭ管理フラグ５１のＶを“１”にセットする（ステップ３０２）。さらに、揮発キャッシュメモリ１３からディスクドライブ８上に書き込みデータを転送した後、ＣＭ管理フラグ５１のＡを“１”にセットする（ステップ３０３）。最後に、ＮＶＳ管理フラグ５０のV_Nを“１”にセットする（ステップ３０４）。この一連の操作は任意契機で実行可能である。

【００５９】一方、任意の契機で発生するホスト側からのリード要求の処理は、一例として、図９に例示されるフローチャートのようにして行われる。

【００６０】すなわち、リード要求が発生すると、まず該当する揮発キャッシュメモリ１３のＣＭ管理フラグ５１がチェックされ（ステップ４０１）、Ａ＝１かつＶ＝０の場合には、リード要求されたデータに対応した未反映の書き込みデータが不揮発キャッシュメモリ１２に存在すると判定して、不揮発キャッシュメモリ１２からデータを読出してホストに転送する（ステップ４０４）。

【００６１】また、ステップ４０１において、Ａ＝１かつＶ＝０でないと判定された場合には、さらに、Ａ＝０かつＶ＝１、または、Ａ＝１かつＶ＝１か否かを調べ（ステップ４０２）、この条件が成立する場合には、揮発キャッシュメモリ１３のキャッシュヒットとして、揮発キャッシュメモリ１３内のデータを読出してホスト側に転送する（ステップ４０５）。

【００６２】ステップ４０１、ステップ４０２のいずれの条件にも合致しない場合には、キャッシュミスと判定し、ディスクドライブ８からデータを読み出し、揮発キャッシュメモリ１３に書き込みつつ、ホスト側にデータを転送し、ＣＭ管理フラグ５１のＡおよびＶを“１”にセットする（ステップ４０３）。

【００６３】このようなＮＶＳ管理フラグ５０およびＣＭ管理フラグ５１を用いた一連の処理により、データ書き込み要求に際してのデータ書き込み動作が、不揮発キャッシュメモリ１２以下の揮発キャッシュメモリ１３、さらにディスクドライブ８のどのレベルで未実行であるか否かに関係なく、ホスト側からのリード要求に対して、最新データのリードを的確に実行可能であり、たとえば、最新の書き込みデータが未反映の古いデータを誤って読出してホスト側に転送する、等の障害の発生を確実に回避することができる。

【００６４】また、このような管理に際してアクセスされるデータは、高々数ビットであるため、ＮＶＳ管理フラグ５０およびＣＭ管理フラグ５１の操作に起因するオーバーヘッドはリード／ライト処理のスループットにはほとんど影響しない。

【００６５】以上説明したように、本実施の形態のキャッシュメモリの配置方法およびデータ記憶システムによれば、複数のチャネルＩ／Ｆ４やパス２を備えた多チャネルパス化、ディスクドライブユニット１０におけるディスクドライブ８の数量の増大によって多ディスクドライブ化されたディスク制御ユニット９の構成において、揮発キャッシュメモリ１３をいくつかの経路毎に分散配置することで、高スループット化が可能となり、さらに、揮発キャッシュメモリ１３とは別個に不揮発キャッシュメモリ１２を集中的に配置して管理することで、不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３の各々のサイズを最適に設定することができ、実装面で有利に、また低価格で高性能なディスク制御ユニット９、すなわち、ディスクサブシステムを実現することが可能となる。

【００６６】なお、不揮発キャッシュメモリおよび揮発キャッシュメモリの分散配置方法としては、図１に例示された方法に限らず、たとえば、図１０～図１２に例示された構成を用いることもできる。なお、図１０～図１２において図１と共通な構成要素には共通の符号を付して説明は割愛する。

【００６７】すなわち、図１０の場合には、ホスト側の複数のチャネルＩ／Ｆ制御回路１ａ～１ｄと、ディスクドライブ８側の複数のディスク制御回路７ａ～７ｄとが、別個に配置される不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３を介して接続される構成としたものである。このような構成においても、上述の図１に例示される構成における効果とともに、揮発キャッシュメモリ１３の制御回路をより簡略化できる、という利点がある。

【００６８】図１１の場合は、ホスト側の複数のチャネルＩ／Ｆ制御回路１ａ～１ｄと、ディスクドライブ８側の複数のディスク制御回路７ａ～７ｆとの間をデータパススイッチ１１を介して接続した構成において、ディスクドライブ８毎に系列をなす、複数のディスク制御回路７ａ，７ｂ，７ｃ，７ｄ，７ｅ，７ｆ，の各系列毎に、互いに独立な不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３の組を配置したものである。この図１１の構成の場合には、ディスクドライブ８の系列毎

に、不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３の組み合わせにおける容量やスループットの組み合わせの最適化を実現できる、という利点がある。

【００６９】図１２の場合には、図１１におけるデータパススイッチ１１を省略するとともに、ディスクドライブ８の各系列が、いわゆるＲＡＩＤにおけるパリティグループを構成し、各パリティグループ毎に、互いに独立な不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３の組を分散して配置したものである。この場合には、たとえば各パリティグループ毎に稼働状況が異なる場合に、当該各パリティグループ毎の不揮発キャッシュメモリ１２および揮発キャッシュメモリ１３の組み合わせにおける容量やスループットの組み合わせの最適化を実現できる、という利点がある。

【００７０】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【００７１】たとえば、データ記憶システムとしてはディスクサブシステムに限らず、記憶階層を有する一般のデータ記憶システムに広く適用することができる。

【００７２】

【発明の効果】本発明のキャッシュメモリの配置方法によれば、キャッシュメモリを経由したデータ転送のスループットを向上させることができる、という効果が得られる。

【００７３】また、本発明のキャッシュメモリの配置方法によれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおけるコスト削減を実現することができる、という効果が得られる。

【００７４】また、本発明のキャッシュメモリの配置方法によれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおける実装効率の向上を実現することができる、という効果が得られる。

【００７５】また、本発明のデータ記憶システムによれば、キャッシュメモリを経由したデータ転送のスループットを向上させることができる、という効果が得られる。

【００７６】また、本発明のデータ記憶システムによれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおけるコスト削減を実現することができる、という効果が得られる。

【００７７】また、本発明のデータ記憶システムによれば、不揮発キャッシュメモリと揮発キャッシュメモリとが混在する構成のキャッシュメモリにおける実装効率の向上を実現することができる、という効果が得られる。

【図面の簡単な説明】

【図１】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの構成の一例を示す概念図である。

【図２】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるリードコマンドの実行時のデータフローの一例を示した概念図である。

【図３】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるライトコマンドの実行時のデータフローの一例を示した概念図である。

【図４】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおける複数のリード／ライト発生時のデータフローの一例を示す概念図である。

【図５】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおいて、複数のリード／ライトが特定のディスクドライブに集中した場合のデータフローの一例を示す概念図である。

【図６】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおいて用いられる制御情報の一例を示す説明図である。

【図７】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるデータ書き込み処理の一例を示すフローチャートである。

【図８】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるデータ書き込み処理の一例を示すフローチャートである。

【図９】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムにおけるデータ読み出し処理の一例を示すフローチャートである。

【図１０】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの変形例を示す概念図である。

【図１１】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの変形例を示す概念図である。

【図１２】本発明のキャッシュメモリの配置方法が実施されるデータ記憶サブシステムの変形例を示す概念図である。

【図１３】考えられる従来の、ディスク制御装置および配下のディスクドライブユニットからなるディスクサブシステムの構成の一例を示す概念図である。

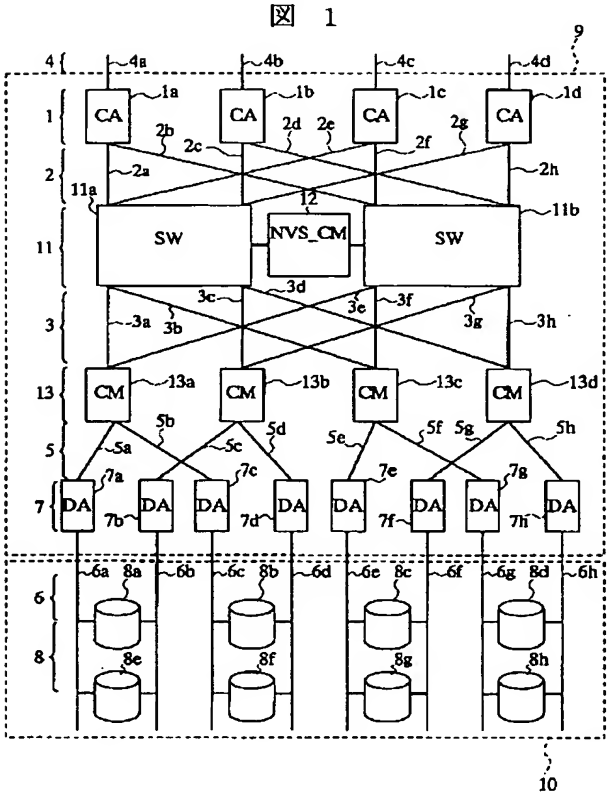
【符号の説明】

１（１ａ～１ｄ）…チャネルＩ／Ｆ制御回路、２（２ａ～２ｈ）…パス、３（３ａ～３ｈ）…パス、４（４ａ～４ｄ）…チャネルＩ／Ｆ、５（５ａ～５ｈ）…パス、６（６ａ～６ｈ）…パス、７（７ａ～７ｈ）…ディスク制御回路、８（８ａ～８ｈ）…ディスクドライブ、９…ディスク制御ユニット（記憶制御装置）、１０…ディスクドライブユニット（記憶装置）、１１（１１ａ，１１ｂ）…データパススイッチ、１２…不揮発キャッシュメモリ、１３（１３ａ～１３ｄ）…揮発キャッシュメモリ、１６，１７…データフロー、２１～２３…データフ

ロー、３４～３７…データフロー、４３～４６…データフロー、５０…ＮＶＳ管理フラグ、５１…ＣＭ管理フラグ。

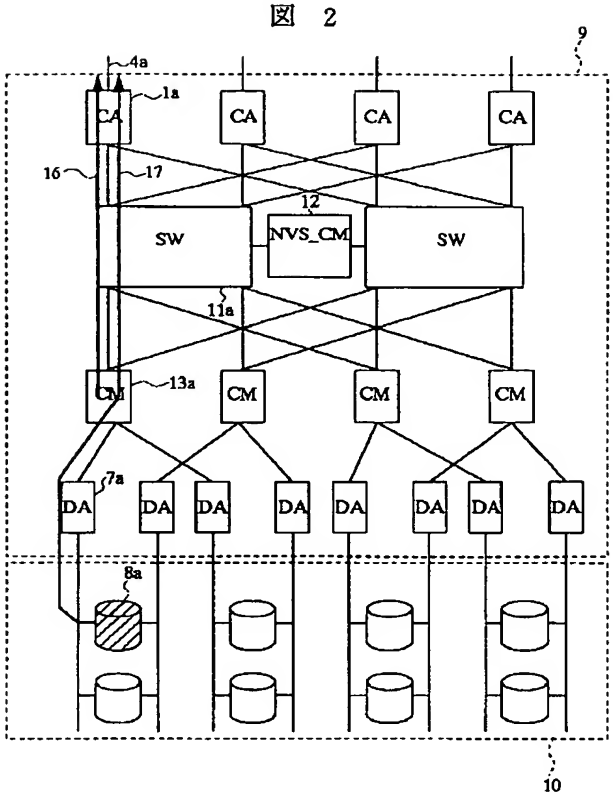
【図１】

図　1



【図２】

図　2



【図６】

図　6

NVS管理フラグ　| V_N |

$V_N = 0$　：　ＮＶＳ上の書き込みデータはＣＭに未反映。

$V_N = 1$　：　ＮＶＳ上の書き込みデータはＣＭに反映済。

CM管理フラグ　| V | A |

Ｖ＝０　：　ＮＶＳからＣＭに未反映の書き込みデータあり。

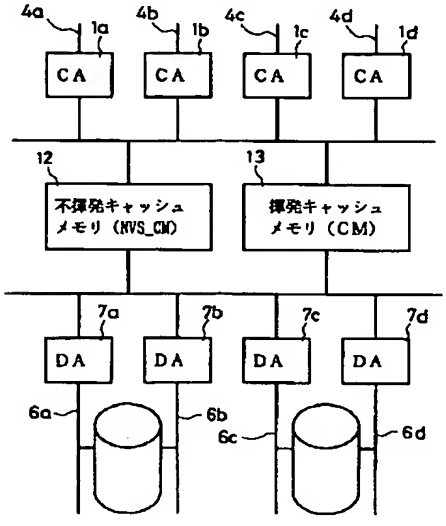
Ｖ＝１　：　ＮＶＳからＣＭに未反映の書き込みデータなし。

Ａ＝０　：　ＣＭからＤＩＳＫに未反映の書き込みデータあり。

Ａ＝１　：　ＣＭからＤＩＳＫに未反映の書き込みデータなし。

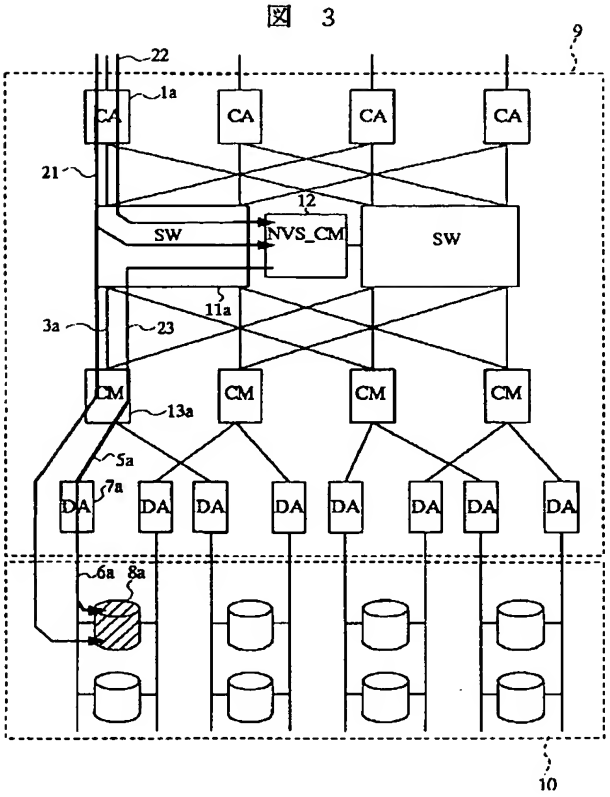
【図１０】

図　10



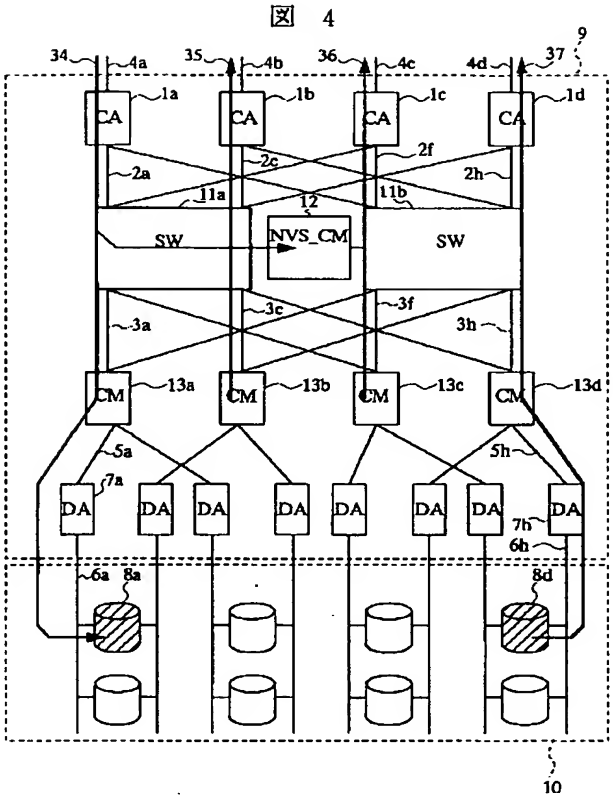
【図３】

図 3



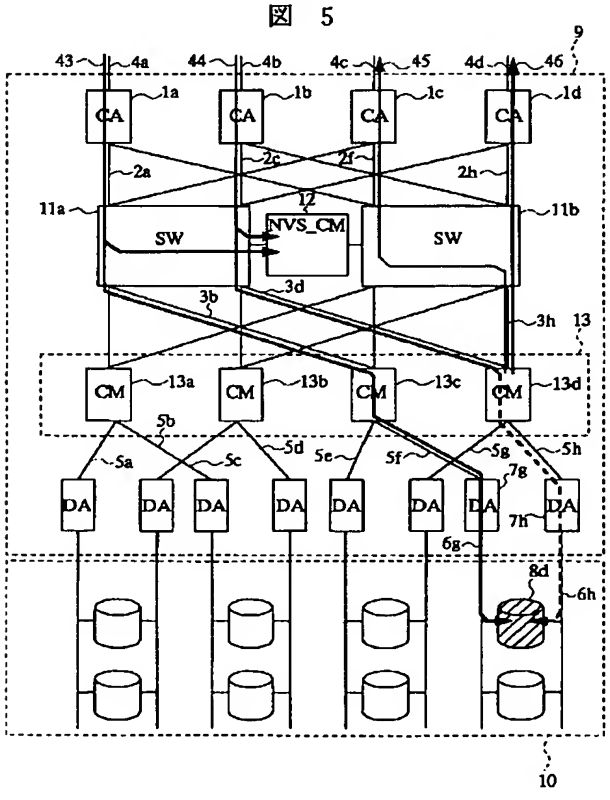
【図４】

図 4



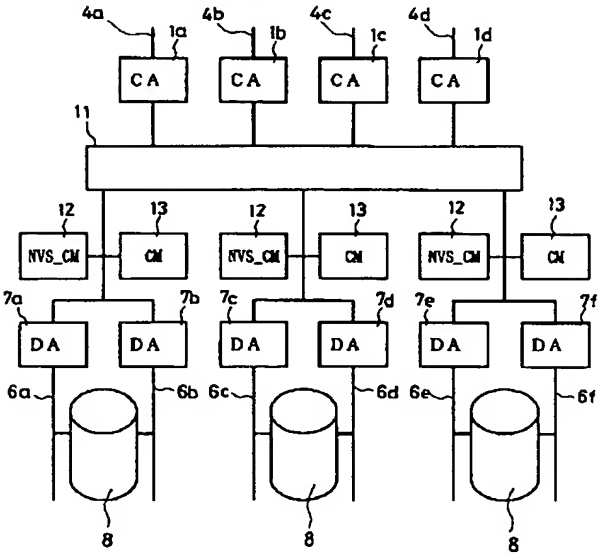
【図５】

図 5



【図１１】

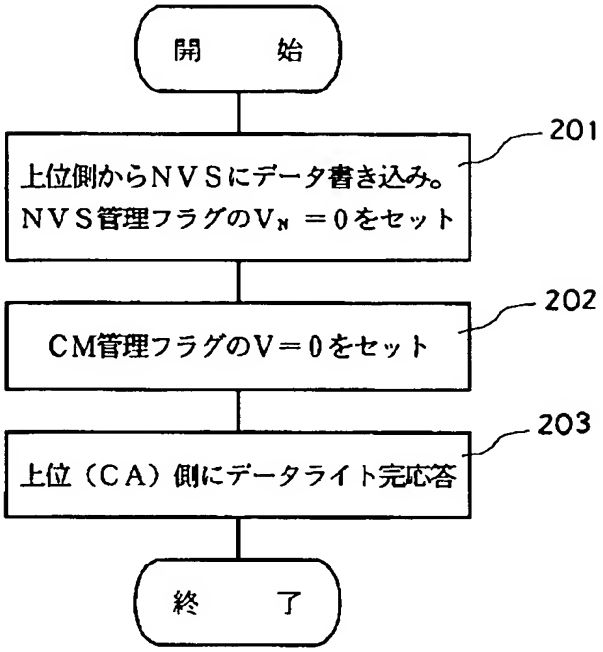
図 11



【図７】

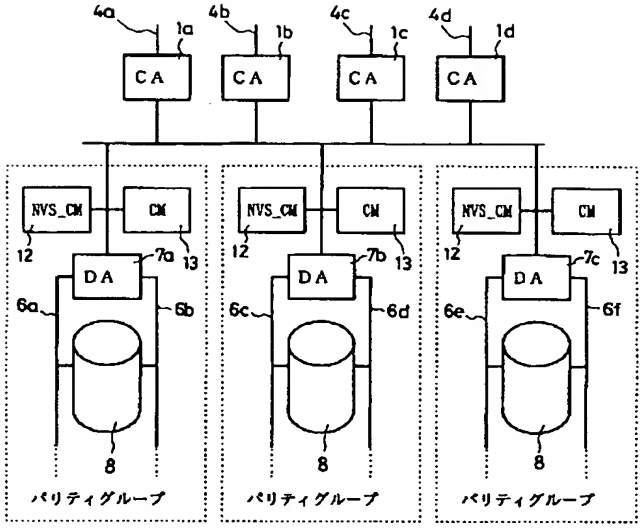
図 7

上位（ＣＡ）側からのデータ書き込み要求処理



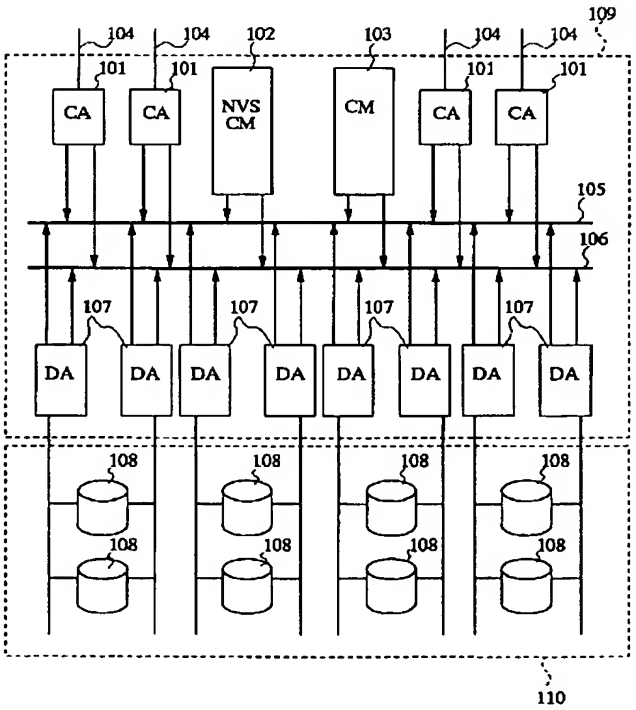
【図１２】

図 12



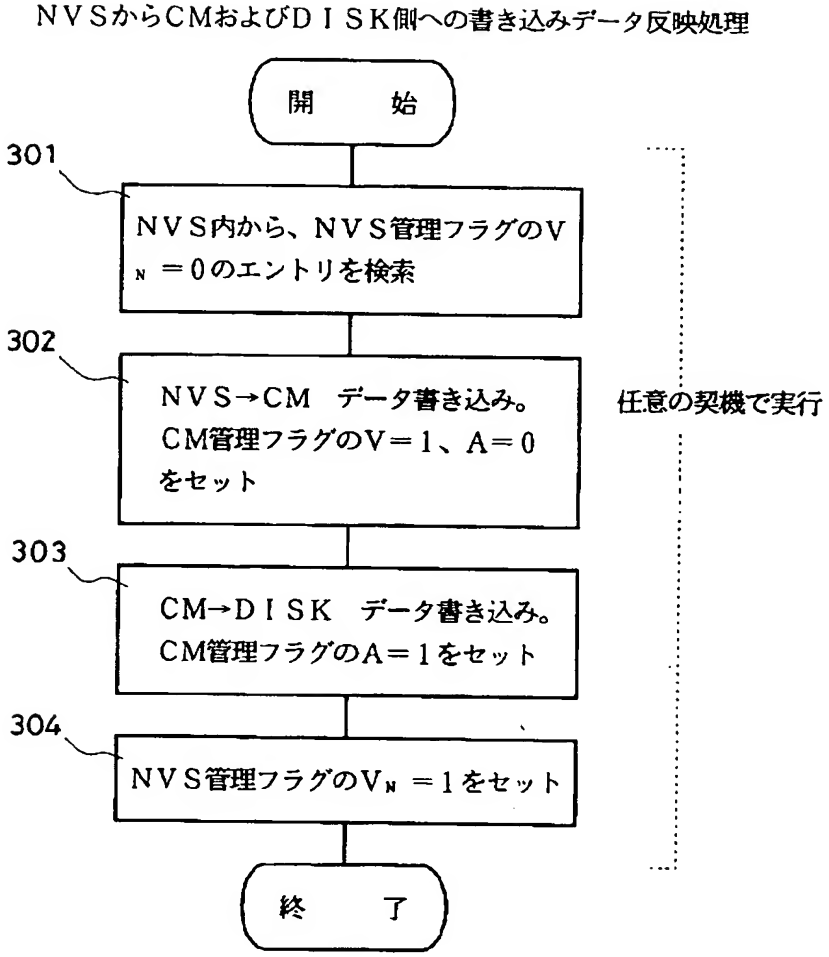
【図１３】

図 13



【図８】

図 8


ＮＶＳからＣＭおよびＤＩＳＫ側への書き込みデータ反映処理
開　始
301
ＮＶＳ内から、ＮＶＳ管理フラグのＶＮ＝０のエントリを検索
302
ＮＶＳ→ＣＭ　データ書き込み。
ＣＭ管理フラグのＶ＝１、Ａ＝０
をセット
303
ＣＭ→ＤＩＳＫ　データ書き込み。
ＣＭ管理フラグのＡ＝１をセット
304
ＮＶＳ管理フラグのＶＮ＝１をセット
終　了
任意の契機で実行

【図９】

図　9

上位(CA)側からのデータ読み出し要求処理